DOCUMENT RESUME

ED 080 570 TM 003 086

TITLE	Pilot Program on Common Status Measures Objective-Referenced Tests. Colorado Evaluation Project, Report No. 1.
INSTITUTION	Colorado State Dept. of Education, Denver.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	27 Sep 70
CONTRACT	OEC-0-70-3731-(284)
NOTE	40p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; *Achievement Tests; *Comparative Analysis; Educational Needs; Educational Objectives; Evaluation Techniques; Item Analysis; *Performance Factors; Public Schools; School Districts; *State Programs; *Test Results
IDENTIFIERS	*Colorado

ABSTRACT

The purpose of the Colorado Evaluation Project was to field test the Common Status Measures at grades four and eleven in conjunction with a statewide assessment program based on objective-referenced testing instruments developed by the Colorado Department of Education for grades kindergarten, three, six, nine, and twelve. The evaluation was designed to determine educational needs and educational program performance in Colorado. Multiple-matrix sampling techniques were used. Results of several types of analyses are presented: (1) for each item and item pool, performance statistics for the total pupil sample and by sub-categories, including school program participation, urbanism, sex, ethnic background, second language spoken at home, and socioeconomic status; (2) a profile of those students in the lower 20% on each subject tested for a given grade level; (3) performance statistics for each item and item pool for each school and district; (4) comparisons between national data for the Common Status Measures; (5) item quality, objective appropriateness and student performance for each objective-referenced item pool; (6) item intercorrelations for each subject at each grade level; (7) subject-by-subject correlations within each grade level and Common Status Measures by subject correlations for adjacent grade levels; and (8) production of two tapes, one concerning occupational cognizance at grade 11 and one concerning language arts in grade 3. (For related documents, see TM 003 087, 088.) (Author/KM)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

67
00203
SP
CG

COLORADO DEPARTMENT OF EDUCATION
Denver, Colorado

September 27, 1970

COLORADO EVALUATION PROJECT

Report Number One

PILOT PROGRAM

on

Common Status Measures
Objective-Referenced Tests

Arthur R. Olson, Director
Assessment and Evaluation
Colorado Department of Education
Denver, Colorado

John W. Helper, Assistant Project Director
Colorado Evaluation Project
Denver, Colorado

Robert W. Heath, Director
Pacific Educational Evaluation Systems
Stanford, California

TM 003 086

Financed by a grant to the Colorado Department of Education from the United States Office of Education under Title IV of the Elementary and Secondary Education Act of 1965 (P. L. 89-10), Pacific Educational Evaluation Systems, Sub-Contractor, Contract No. OEC-0-70-3731 (284)

Byron W. Hansford
Commissioner of Education

Table of Contents

Page

INTRODUCTION.....1

The Purpose.....1

Genesis.....1

Rationale.....2

Working Definitions.....2

"Objective".....2

"Achievement".....3

"Anchor".....4

"Educational need".....4

Procedures for Statewide Testing.....4

Sampling.....4

Hiring proctors.....4

Analysis.....5

FINDINGS.....5

The Common Status Measures.....5

Occupational cognizance.....6

Verbal status.....8

Common Status Measures and Needs Assessment.....10

Objective-referenced Items.....13

An example.....13

Objectives.....14

Student groups.....16

Items.....24

Other Comparisons.....24

CSN and objective-referenced correlations.....27

Teacher judgment and pupil achievement.....27

Page

Summary.....27

CONCLUSIONS.....29

RECOMMENDATIONS FOR ACCOUNTABILITY.....30

Stating Objectives.....30

Developing Test Items.....30

Testing.....31

Analysis.....31

Recycling.....31

RELATED DOCUMENTS.....32

THEORETICAL WORKS.....32

APPENDIX A: PUPIL CHARACTERISTIC ITEMS.....33

APPENDIX B: CONTENT AREAS - SAMPLED BY OBJECTIVES-REFERENCED ITEMS.....35

List of Tables and Graphs

Table		Page
1	Number of Objectives Written by Grade Levels and Content Areas	3
2	Estimated Percent Correct for All Students on All Items: Colorado Sample and National Sample on The Common Status Measures	6
3	Comparison of National with Colorado Scores on Elements of Occupational Knowledge	7
4	Common Status Measure Total Item Pool Results - Occupational	9
5	Common Status Measure Total Item Pool Results - Verbal	11
6	Description by Classification Groups of the Composition of the Lowest 20% Performance Group by Test Within Grade Level	18
7	Objective-Referenced Total Item Pool Results - Kindergarten	19
8	Objective-Referenced Total Item Pool Results - Grade Three	20
9	Objective-Referenced Total Item Pool Results - Grade Six	21
10	Objective-Referenced Total Item Pool Results - Grade Nine	22
11	Objective-Referenced Total Item Pool Results - Grade Twelve	23
12	Third - Grade Subject and Fourth - Grade Common Status Measure Spearman Rank Order Correlation (Rho)	25
13	Eleventh - Grade Common Status Measure and Twelfth - Grade Subject Spearman Rank Order Correlation (Rho)	25

Graph		
1	Responses to Objectives: Teacher Judgment, Pupil Achievement	26

INTRODUCTION

The Colorado Evaluation Project was initiated to field test the Common Status Measures (CSM) concurrently with Colorado's pilot program in assessment and evaluation. The project's successful completion was possible only with the cooperation of many persons. Dr. Gene Glass of the University of Colorado advised the Colorado Department of Education (CDE) on item development for objective-referenced items. The Belmont Group supplied interest and support throughout the entire project. United States Office of Education personnel were involved in many aspects of the project. Most importantly, the educators and school children of Colorado gave the effort and time required to make the project worthwhile.

The Purpose

The primary purpose of the Colorado Evaluation Project was to determine Colorado's educational need, defined as "the discrepancy between stated objectives and their achievement." A secondary purpose was to test assessment procedures that may be replicated or adapted by local school districts, by other states, and by federal agencies.

Genesis

The interest of the Colorado Department of Education in precise assessment was the product of two great current streams of concern and thought, and was coincidental in timing with the emergence of a third great stream. The over-arching questions arising from these streams were these:

How well do Colorado students learn?

What are strengths and weaknesses?

If we want improvement, how do we get it?

These questions are both scientific and political; scientific in the realms of testing, measurement, sampling and computer analysis - political in the realms of allocation of local, federal and state resources. Science and politics constitute the streams of concern and thought mentioned above.

Rationale

The confluence of scientific and political inquiry has produced a third stream and a new term in the schools and legislative halls of our land: "Educational Accountability". Properly understood, accountability can lead us all to an awareness of common responsibility for the education of children in Colorado. Our efforts and our resources are expended for quality education. To this end Colorado educators, legislators, and others are all accountable.

Working Definitions

Throughout the project, the investigators were guided by the definition of "educational need" as "the discrepancy between objectives and their achievement". This definition, in its simplicity, entailed further explication as the project developed. These explications follow.

"Objective." Specifications for objectives shaped their definition. Colorado educators decided that each objective specify what a child should be able to do upon completion of a given year in school in terms of (1) the subject-content area (2) the action or performance desired, and (3) the measuring instrument or technique. Further, each goal was to be consistent with Goals for Education in Colorado, as adopted in 1962 and restated in 1970 by the Colorado State Board of Education. Finally, CDE personnel

specified that each objective be considered important by Colorado teachers. Accordingly, each objective was judged by a random sample of teachers in Colorado as to their importance or unimportance as year-end objectives. Level of importance was indicated by per cent of teachers so responding.

Numbers of objectives so written and judged are indicated in Table 1 below.

TABLE 1

Number of Objectives Written

by Grade Levels and Content Areas

Level	Health	P.E.	Math	Lang. Arts	Science	Music	Total
K			8				22
3	28	12	10	21	12	55	138
6	27	15	17	14	10	60	143
9	17	26	16	22		49	130
12	15	26	14	21	—	49	125
Totals	87	79	65	78	36	213	558

"Achievement." Achievement was defined in terms of percentage of correct responses to test items relating to the objectives described above. Items were developed and inspected for face validity with the basic question - "Does the item really measure what it appears to measure?" Reading level, hidden meanings, ambiguities, and other faults were corrected as time permitted. Logical relationships of items to objectives were evaluated during item development and later data analysis. The items did not have to fix any individual's position with respect to norms, nor predict his further success, nor correlate with other abilities, nor be easy

for some and hard for others. The items simply had to reveal whether the individual could or could not perform a given task.

"Anchor." The anchor concept developed rather late in the project and is explained here. The Common Status Measures (CSM) provided an anchor for the Colorado Evaluation Project. By releting national scores obtained previously on the CSM with scores obtained in the Colorado project, new data could be compared with existing data to determine stability of the tests and testing procedures used in Colorado.

"Educational need." Derived from the above definitions, "educational need" was a construct with several facets. The need was felt by teachers who judged educational objectives, by curriculum specialists who wrote the objectives, and by pupils who were puzzled by some of the test items. Such discrepancies have been described in educational literature as the difference between the "is" and "ought", between the "real" and "ideal", between "balance" and "imbalance" and many other theoretical constructs dealing with discrepancies.

Procedures for Statewide Testing

Multi-matrix sampling, hired test proctors, and computer analysis were basic procedures utilized for the Colorado Evaluation Project. Some 12,000 tests were administered in 209 schools in 31 districts across the state. Details are reported in the project's interim report (June 15, 1970) and are summarized briefly below.

Sampling. Multi-matrix sampling provides random selection of both pupils and test items. In this way expenses of every pupil taking every item are avoided while needed information is produced.

Hiring proctors. Test proctors were hired at $20.00 per day

plus mileage to give the tests in the schools. This avoided disruption of ongoing teaching-learning school activities.

Analysis. Responses to test questions were key-punched on data cards. A computer program of the Pacific Educational Evaluation System, Stanford, California, provided analyses comparing pupil responses and pupil characteristics along with numerous other statistical procedures as described in the following section on findings.

FINDINGS

The basic unit of information produced were percentages of correct and incorrect responses to a test item. This type of information was tabulated to facilitate comparisons among various student populations and among the items. Selected statistical operations indicated precision in these comparisons.

Test scores are discussed below on (1) the Common Status Measures (2) the objective-referenced achievement tests, and (3) comparison of results from these two types of instruments.

The Common Status Measures (CSM)

The United States Office of Education contracted in 1969 to develop the CSM, U.S.O.E. Contract EC 0-9-099017-4424(010). Under the U.S.O.E. auspices, two 72-item pools were administered in Colorado, one item pool testing occupational cognizance, the other testing basic verbal status. Both 4th and 11th graders tried items from both pools. Comparisons of the Colorado and national samples are made in the following table in terms of estimated percentage of correct responses of all students taking all items.

TABLE 2

Estimated Percent Correct
for All Students on All Items:
Colorado Sample and National Sample on
The Common Status Measures

Common Status Measure of:	Colorado Sample		National Sample		Difference	
	Grade 4	Grade 11	Grade 4	Grade 11	Grade 4	Grade 11
Basic Verbal Status	74.3	66.1	67.6	59.0	6.7	7.1
Occupational Cognizance	62.0	67.8	47.0	61.0	15.0	6.8

It may be observed that Colorado scores are higher. The scores appear fairly consistent except at the fourth grade level in occupational cognizance.

Occupational cognizance. Four elements comprised occupational cognizance in the Common Status Measures. These elements tested knowledge of a given occupation regarding (1) education or training needed, (2) work involved, (3) other occupations related, and (4) field of work. Colorado and national scores were compared on these four elements.

TABLE 3

Comparison of National with Colorado Scores* on Elements of Occupational Knowledge in 4th and 11th Grades

Elements tested:	National 4th	National 11th	Colorado 4th	Colorado 11th	Average Difference 4th	Average Difference 11th
Education or training needed	36	56	50	65	14	9
Work involved	57	62	71	65	14	3
Occupations related	43	59	64	66	16	7
Field of work corresponding	52	65	69	69	17	4

*Estimated percentage correct for all pupils taking all items

The table indicates that Colorado's edge is larger in the 4th than in the 11th grade. Colorado 11th graders scored higher than the national sample, indicated in the "difference" column at the right. However, the differences achieved by the Colorado 4th graders are considerably higher than those achieved by the Colorado 11th graders.

Perhaps the most useful information regarding needs for occupational cognizence can be found in comparing student groups. Chi-square comparisons showed certain groups to have more than their share of low scorers. This information was most useful in answering questions posed at the outset of the project regarding occupational cognizence in Colorado.

Q: What are the distinguishing characteristics of pupils deficient in knowledge of occupations?

A: At both 4th and 11th grade levels, several student groups had disproportionate difficulty with questions on occupations. Students from Title I schools comprised one low

scoring group; these students come from families with low annual income. Similarly 4th graders from low socio-economic status families had more than their share of difficulty. Also, at both grade levels, American Indian, Negro and Spanish-surnamed American children significantly had more difficulty than the total student population. Eleventh graders who were from bilingual homes or who were girls had more than their share of difficulty with questions on occupations. A single low-scoring student may have several of these characteristics.

Q: Are there significant differences in level of occupational cognizance and basic verbal status with respect to urbanism and location of school?

A: No significant difference in occupational cognizance was found relating to urbanism of the school's district. Urban students tended to outscore either suburban or rural students, but not to a significant extent; such differences may have been due to chance. District characteristics, such as isolation or geographic location were not separated in this analysis.

Q: What are the relations among occupational cognizance and characteristics of school and program?

A: Somewhat surprisingly, fourth graders from schools with Title III programs showed up with disproportionately low scores on occupational cognizance. Title III schools have innovative programs not necessarily related to low achievement. Low results in both fourth and eleventh grades also came in from schools with both Title I and Title III programs.

Table 4 following gives the percentage correct scored on occupational questions. Column headings are explained in the explanatory information above.

Verbal status. Although Colorado pupils outscored the national sample on verbal status, certain student populations within Colorado missed more than their share of those items measuring verbal status. More particularly, however, the data from the Colorado Evaluation Project contain answers to the questions posed at the outset of the project.

Q: What is the distribution of functional literacy over the state for pupils in grade 4 and 11?

TABLE 4

COMMON STATUS MEASURE TOTAL ITEM POOL RESULTS

OCCUPATIONAL

Subject p = % correct	Total	Title No.				Urbanism			Sex		Ethnic Group					2nd Lang.		SES		
		1	3	B	N	U	S	R	M	F	Al	N	O	SA	Other	Y	N	L	M	H
Grade 4 Occupational p	62.0	59	47	52	67	64	56	58	63	61	53	44	57	47	66	54	64	56	63	72
Grade 11 Occupational p	67.8	63	71	64	68	69	62	61	70	65	69	51	*	50	70	58	69	64	69	72

*Sample size less than 5 students

A: Groups of students with more than their share of difficulty on verbal items were these in the 4th grade: Negro, Spanish-surnamed Americans and low socio-economic status groups. Distribution of ethnic minorities in Colorado is described in the recent CDE report Colorado People and Colorado Education. Almost 100% of the Negro population is located in Denver, and comprises between 10 and 15 per cent of that district's school population. The Spanish-surnamed and American Indian students are scattered in rural and urban districts, with lower proportions in the suburbs.

Surprisingly, verbal scores from Title I did not show up significantly low in the fourth grade. In the eleventh grade, scores were significantly low from Title I schools and those schools with both Title I and Title III schools and from pupils who were Negro and Spanish-surnamed American.

Q: What is the relation between literacy level of pupils and characteristics of schools serving them?

A: Interestingly, elementary schools with Title I programs did not send in scores significantly low; Title I high schools did. Schools with both Title I and III programs sent in scores disproportionately low. No relationships were found between low scores and characteristics of urbanism, and suburbanism or rurality.

See Table 5 following for summary of above information.

Common Status Measures and Needs Assessment

It should be remembered that the Common Status Measures are item-pools not tests. That is, at the fourth-grade level we have a collection of 72 questions that sample several aspects of pupils' basic verbal status. It follows then that we should look to performance on these items first when using the CSM's for the assessment of educational needs.

If we continue to use Basic Verbal Status-Fourth Grade as our example, we could look at the performance of different groups of pupils (or schools, districts, states, regions, etc.,) on reading comprehension questions and on vocabulary questions. Each of these questions yields curricular-relevant information. For example, the response to a vocabulary question will not only

TABLE 5*

COMMON STATUS MEASURE TOTAL ITEM POOL RESULTS

VERBAL

Subject p = % correct	Total	Title No.				Urbanism			Sex		Ethnic Group					2nd Lang.		SES		
		1	3	S	N	U	S	R	M	F	AI	N	O.	SA	Other	Y	N	L	M	H
Grade 4 Verbal p	74.3	74	63	69	77	76	70	73	74	75	71	62	70	65	77	60	75	69	73	84
Grade 11 Verbal p	66.1	62	70	60	66	67	64	60	67	65	66	46	*	51	63	58	67	64	67	72

*Sample size less than 5 students

tell us how many of what kind of pupils answered correctly, but also if the incorrect response were "paradigmatic" (work-play), "syntigmatic" (work-hard), "phonologic" (walk-work), based on a spelling similarity (word-word) or based on an opposite (work-relax). This type of fine-grain analysis can be made with each item in the four item pools. We might find-hypothetically-that Spanish-surnamed children in a particular rural area were doing well on most vocabulary items, but consistently poorly on reading comprehension questions. Perhaps an especially bad E.S.L. program that drills on mechanics, but neglects reading comprehension is being used in this area or perhaps the children were learning the elements of English well, but were given only reading materials that had little relevance to their lives and their cultural setting. Whatever the reason, a need had been identified.

The Common Status Measures can, of course, be used in needs assessment at other than the test-question level. For example, it is possible to estimate the over-all performance of a pupil group on an entire item pool. We learned that 11th grade girls in Colorado do poorly in the occupational cognizance item pool. We can probably recognize a historical sex bias in vocational education, and we needn't wait for the women's liberation movement to begin correcting it. Again, a need was identified.

The Common Status Measures are a little like the physician's x-rays. They can point out anomolies, failures, triumphs, and normalcy. They do not substitute for the professional skill and judgement of the practitioner. To be used

effectively, the Common Status Measures must be interpreted by educators close enough to the schools to know what implications the results have in the context of the educational system.

Objective-referenced Items

Scores on the tests were reported by percentages of correct and incorrect responses to each test item. Further analyses produced comparisons among various pupil populations and among the many test items. Performance scores were compared among these populations:

ESSA Title No: I, III, Both I & III (B) and Neither (N)

Urbanism: Urban (U), Suburban (S) and Small (R)

Sex: Male (M) Female (F)

Ethnic Group: American Indian (AI) Negro (N) Oriental (O) Spanish-surnamed American (SA) and other,

Bilingual Home: Yes (Y) No (N)

Socio-economic Status (SES): Low (L) Middle (M) High (H)

(The letters in parentheses refer to column headings on Tables 4 through 11.)

Information was reported regarding (1) the objectives upon which the items were based (2) Colorado students, and (3) the test items produced by Colorado educators. This information is summarized below following an example illustrating the development of the objective-referenced tests in Colorado.

An example. This example of objective-referenced testing shows its development and use. A goal for education in Colorado,

as authorized by the State Board of Education, states as follows:

> Goal: Adequate opportunities for all persons to acquire command knowledge, skills, habits and attitudes essential for effective learning throughout life. (Goals for Education in Colorado, 1962)

From this goal, Colorado educators derived this objective for 9th and 12th grade language arts:

> Objective: The pupil applies independent research methods by preparing and presenting oral and written information on a topic as measured by a check list that includes the following skills: Note-taking, summarizing, information gathering from more than one source, footnoting, quoting, citing references, drawing inferences, organizing and outlining.

Fifty-eight per cent of the ninth graders handled those test items well, and the summary comment of the first evaluators was, "Student performance indicates a need for improvement in this area." But 77 per cent of the twelfth graders were correct in these test items, and the summary comment was, "Student performance indicates that students have a good mastery of this objective."

Thus, somewhere between ninth and twelfth grade, Colorado students acquire a respectable mastery of this skill. Here Colorado educators have a specific opportunity to take a look at something that's being done right and well, and from it learn something that should help them with other skills that are not being taught so well.

Objectives. Colorado teachers judged importance of curricular objectives on which the testing was based. The teachers judged some objectives to be "very important" while judging

others to be not so important. The "$64 question" was ARE OBJECTIVES JUDGED TO BE IMPORTANT BY COLORADO TEACHERS BEING ACHIEVED BY COLORADO STUDENTS? The discrepancy between an objective and its achievement defined an educational need.

On one of the objective-referenced test items for third-grade mathematics, pupils were to display their knowledge of money by identifying change in coins with purchases up to $5. Only 50 per cent of the third graders tested answered one of the test items correctly, and only 28 per cent answered the other item correctly. The conclusions reached by the examining educators was that the objective was a reasonable one for third grade and that the test item fairly tested the knowledge; so it is evident that something is wrong in the teaching-learning equation, for the pupils simply are not doing well enough. What is to be done as a result is not yet defined. This will require the insight and experience of curriculum experts, of mathematics experts, of successful teachers. There will be a reaction, and a proposal for change; and if it works, in a succeeding test round, Colorado third graders will show more skill in handling change from a $5 bill.

Quite a different reaction occurred with a third-grade question in science, probing the pupil's understanding of the relationship of position and motion. Only 7 per cent of the third graders answered that one correctly. The first judgment was that the problem lay with one of two things: Either it is overly optimistic to expect a third grader to know such things (and therefore the objective is fanciful) of the test question was poorly drawn (and therefore pupils could not answer that, even though they knew something about the area being tested). Again, more work will be required to determine whether this question should be dropped, or emended.

Student groups. Comparisons among student groups revealed significant differences on percentages of correct and incorrect responses. These comparisons are summarized on Table 6 and described in narrative form according to grade level below. Tables 7 to 11 summarize by grades the percentage of correct responses to total pools of objective-referenced items in the several content areas. (See abbreviations on page 13)

Kindergarten:

Low SES children tended to score lower on the Mathematics items than middle or high SES children. No other differences between groups are significant.

Third Grade:

The Music questions did not discriminate between groups. The Mathematics items showed ethnic group differences. Children who spoke a language other than English in their homes found these questions more difficult than children for whom English was the primary language in the home.

On the Physical Education items children in Title I and III scored lower than other children. All minority ethnic groups found these items difficult.

Title I and B groups students tended to rank lower than other children on the Language Arts item pool. American Indians, Negroes, and Orientals tended to rank lower than other children.

On the Health test, Title III and B groups ranked lower overall than other Title I or N groups. Negroes and Spanish-surnamed Americans tended to score lower than other children as did low SES children.

Sixth Grade

Only urban and low SES children tended to rank low on the Physical Education items at this level.

Math proved relatively more difficult for Title III and B students and for students in suburban and rural schools than for other children.

Language Arts was more difficult for boys than for girls. All ethnic groups except "Other" ranked low as did the low SES group. A second language spoken in the home was also characteristic of low achievement on this test.

The Science test tended to be more difficult for Title III and B group children than for other children. American Indians and Negroes

ranked lower than other ethnic groups.

Girls and most ethnic groups (AI, N, SA) experienced difficulty with the Health questions at this grade level.

Music proved relatively more difficult for Title I, III, and B groups than other children. Rural school children tended to rank lower than other children on this subject as did Spanish-surnamed students, low SES students and bilingual students.

Ninth Grade:

No group was distinguished from the total group on the Health test at this grade level.

On the Math test, Title I and B group children tended to rank lower than other children in that classification. Negroes and Spanish-surnamed Americans found this subject difficult as did children who come from homes where a second language is spoken.

Boys scored higher than girls on the P.E. items.

Language Arts was relatively more difficult for AI, N, O. and SA ethnic groups than for "Others". A second language spoken in the home was also a characteristic of low performance.

Rural school children and boys tended to rank low on Music in the ninth grade. Bilingualism was also a significant characteristic of students with poor performance in Music.

Twelfth Grade:

As at the ninth-grade level, no significant group differences on the Health test are apparent.

Girls did not perform as well as boys on the Physical Education measures.

The Math questions resulted in lower rankings for Title I and B children than for other children. Suburban and rural children also experienced difficulty on this measure, as did Negro and SA ethnic groups. Low SES and second language spoken in the home were more characteristic of the lower twenty percent of students in Math at this grade level than of the upper eighty percent.

Language Arts was relatively more difficult for children in schools under both Title I and III than for other children. Low SES children, suburban children, and children in all minority ethnic groups tended to rank low on this subject.

The Music questions discriminated in all classifications except sex. Title I and III groups did poorly as did suburban school and rural school children. Low SES was characteristic of poor performance.

Students from homes that spoke a second language experienced difficulty with this subject as did Negro and Spanish-surnamed Americans.

Table 6

Description by Classification Groups of the Composition of the Lowest 20% Performance Group by Test Within Grade Level*

		Title				Urbanism			Sex		Ethnic Group					2nd Lang.		SES		
		I	III	B	N	U	S	R	M	F	AI	N	O	SA	OTHER	YES	NO	L	M	H
Kindergarten	Math																	x		
	Science																			
Third Grade	Math										x	x		x		x				
	Phys. Ed.		x	x							x	x	x	x						
	Lang. Arts	x		x							x	x	x							
	Science															x		x		
	Health		x	x								x		x				x		
	Music																			
Fourth Grade	CSM Verbal		x	x								x		x				x		
	CSM Occ.	x	x	x							x	x		x				x		
Sixth Grade	Math		x	x			x	x												
	Phys. Ed.					x												x		
	Lang. Arts								x		x	x	x	x		x		x		
	Science		x	x							x	x								
	Health								x		x	x		x						
	Music	x	x	x				x						x		x				
Ninth Grade	Math	x		x								x		x		x				
	Phys. Ed.									x										
	Lang. Arts										x	x	x	x		x				
	Health																			
	Music							x	x							x				
Eleventh Grade	CSM Verbal	x			x							x		x						
	CSM Occ.	x			x	x				x	x	x		x		x				
Twelfth Grade	Math	x			x		x	x				x		x		x				
	Phys. Ed.									x										
	Lang. Arts				x		x				x	x	x	x				x		
	Health																			
	Music	x	x				x	x				x		x		x		x		

*An "x" indicates that the category has an disproportionately large number of pupils in the lowest fifth of the performance range (significant at the .05 level).

TABLE 7

OBJECTIVE-REFERENCED TOTAL ITEM POOL RESULTS

KINDERGARTEN

Subject p = % correct	Total	Title No. 1	Title No. 3	Title No. B	Title No. N	Urbanism U	Urbanism S	Urbanism R	Sex M	Sex F	Ethnic Group AI	Ethnic Group N	Ethnic Group O	Ethnic Group SA	Ethnic Group Other	2nd Lang. Y	2nd Lang. N	SES L	SES M	SES H
Math p	68.5	74	58	69	70	67	64	76	71	66	*	57	*	68	71	64	68	62	78	63
Science p	65.5	70	57	68	66	65	69	65	65	66	*	61	*	63	68	64	66	69	67	64

*For Tables 3-7, denotes sample size was less than five students

TABLE 8

OBJECTIVE-REFERENCED TOTAL ITEM POOL RESULTS

GRADE THREE

Subject p = % correct	Total	Title No. 1	Title No. 3	Title No. 8	Title No. N	Urbanism U	Urbanism S	Urbanism R	Sex M	Sex F	Ethnic Group AI	Ethnic Group N	Ethnic Group O	Ethnic Group SA	Ethnic Group Other	2nd Lang. Y	2nd Lang. N	SES L	SES M	SES H
Lang. Arts p	69.4	67	66	63	71	68	75	66	68	71	64	52	53	67	72	68	70	67	72	72
Math p	58.6	56	56	51	61	60	53	55	60	57	51	52	*	52	63	54	61	55	62	66
Music p	41.3	41	39	38	42	41	43	45	42	41	34	40	*	37	43	37	43	[illegible]	43	45
P.E. p	59.0	57	42	49	62	60	57	57	59	59	53	43	38	52	62	56	60	[illegible]	62	56
Health p	78.7	80	76	73	80	78	81	79	77	81	78	67	77	73	81	76	80	75	80	82
Science p	52.2	50	46	54	53	51	53	53	52	53	41	45	*	43	55	43	55	45	57	56

* = Sample size less than 5 students

TABLE 9

OBJECTIVE-REFERENCED TOTAL ITEM POOL RESULTS

GRADE SIX

Subject p = % correct	Total	Title No. 1	3	6	8	Urbanism U	S	R	Sex M	F	Ethnic Group AI	N	O	SA	Other	2nd Lang. Y	N	L	SES M	H
Lang. Arts p	65.2	63	62	50	67	66	63	66	63	67	61	58	50	58	67	58	67	60	67	69
Math p	46.5	49	42	34	48	47	47	43	45	48	48	35	52	40	48	42	48	42	48	51
Music p	36.8	35	35	34	38	37	35	32	36	38	34	34	*	31	38	31	37	32	38	43
P.E. p	63.9	46	63	63	64	62	67	72	67	60	64	62	*	62	67	61	65	58	66	69
Health p	66.1	68	63	58	67	66	65	68	64	68	63	58	*	62	68	63	67	63	68	70
Science p	37.6	38	35	34	38	38	37	36	37	38	41	27	50	35	38	36	38	37	38	38

*sample size less than 5 students

TABLE 10

OBJECTIVE-REFERENCED TOTAL ITEM POOL RESULTS

GRADE NINE *

Subject p = % correct	Total	Title No. 1	3	B	N	Urbanism U	S	R	Sex M	F	Ethnic Group AI	N	O	SA	Other	2nd Lang. Y	N	L	SES M	H
Lang. Arts p	71.7	70	79	71	72	71	72	74	71	72	61	65	69	65	74	67	73	73	73	77
Math p	50.6	45	59	46	53	52	46	50	50	51	46	35	64	39	53	41	53	46	53	58
Music p	33.0	30	34	34	34	33	34	29	32	34	30	30	39	29	34	30	34	30	35	35
P.E. p	43.7	43	51	44	44	44	44	44	48	39	56	45	42	40	44	38	45	46	43	46
Health p	78.2	75	85	77	79	78	79	78	77	80	79	75	67	76	80	79	79	77	79	82

TABLE 11

OBJECTIVE-REFERENCED TOTAL ITEM POOL RESULTS

GRADE TWELVE

Subject p = % correct	Total	Title No.				Urbanism			Sex		Ethnic Group					2nd Lang.		SES		
		1	3	B	N	U	S	R	M	F	AI	N	O	SA	Other	Y	N	L	M	H
Lang. Arts p	66.5	70	68	59	71	67	57	69	65	68	57	53	*	54	69	62	67	65	65	71
Math p	61.7	57	66	53	66	63	57	55	64	59	60	50	59	46	64	49	63	52	63	66
Music p	40.4	33	41	36	46	41	36	37	39	41	*	37	*	31	42	33	42	36	42	45
P.E. p	43.7	43	51	44	44	44	44	44	48	39	56	45	42	40	44	38	45	46	43	46
Health p	77.1	79	79	73	79	77	75	77	76	78	73	68	*	71	78	74	77	76	79	80

*sample size less than 5 students

Items. Every question developed by the Colorado educators was based on an explicit statement of an educational objective. As an example of how the results may be used, the PEES staff grouped the items by source-objective and attempted to interpret the implications of the results. The performance of each category of pupil on every item is given on 550 pages of computer printout, available at the Colorado Department of Education. Educational specialists in Colorado, who are closer to the curriculum, will undoubtedly be able to make more detailed interpretations. As one interprets the performance of pupils on an item in relation to the corresponding objective, several possibilities become evident:

1. If the item appeals to be "easy", that is a high percentage of pupils answer it correctly,
   (a) the objective may be successfully taught
   (b) the question may be poorly worded to "give-away" the correct answer
   (c) the objective may have been inappropriate (pupils already had attained the objectives in an earlier grade)
   (d) the question may not correspond to the objective.

2. If the item appears to be difficult,
   (a) the objective may not have been dealt with successfully by the curriculum, teaching methods, or both
   (b) the question may be so poorly worded that pupils chose an incorrect response though they have, in reality, attained the objective
   (c) the objective may be inappropriate (that is, too advanced for the indicated grade level)
   (d) the question may not correspond to the objective

These possibilities were considered in the interpretations made for each of the objective-referenced items.

Other Comparisons

Two over-all comparisons yielded information on relations of (1) the CSM to objective referenced responses and (2) relations of student

Table 12

Third - Grade Subject and Fourth - Grade Common Status Measure Spearman Rank Order Correlation (Rho)*

		Language Arts	Science	Physical Educ.	Health	Music	Math	CSM-V	CSM-OCC
Basic Verbal Status	Rho	.08	.20	.18	.11	.13	.14	---	.69
	N	94	86	86	91	82	89	---	96
Occupational Cognizance	Rho	.19	.33	.32	.27	.06	.31	.69	---
	N	94	86	86	91	82	89	96	---

*Ranking by school mean

Table 13

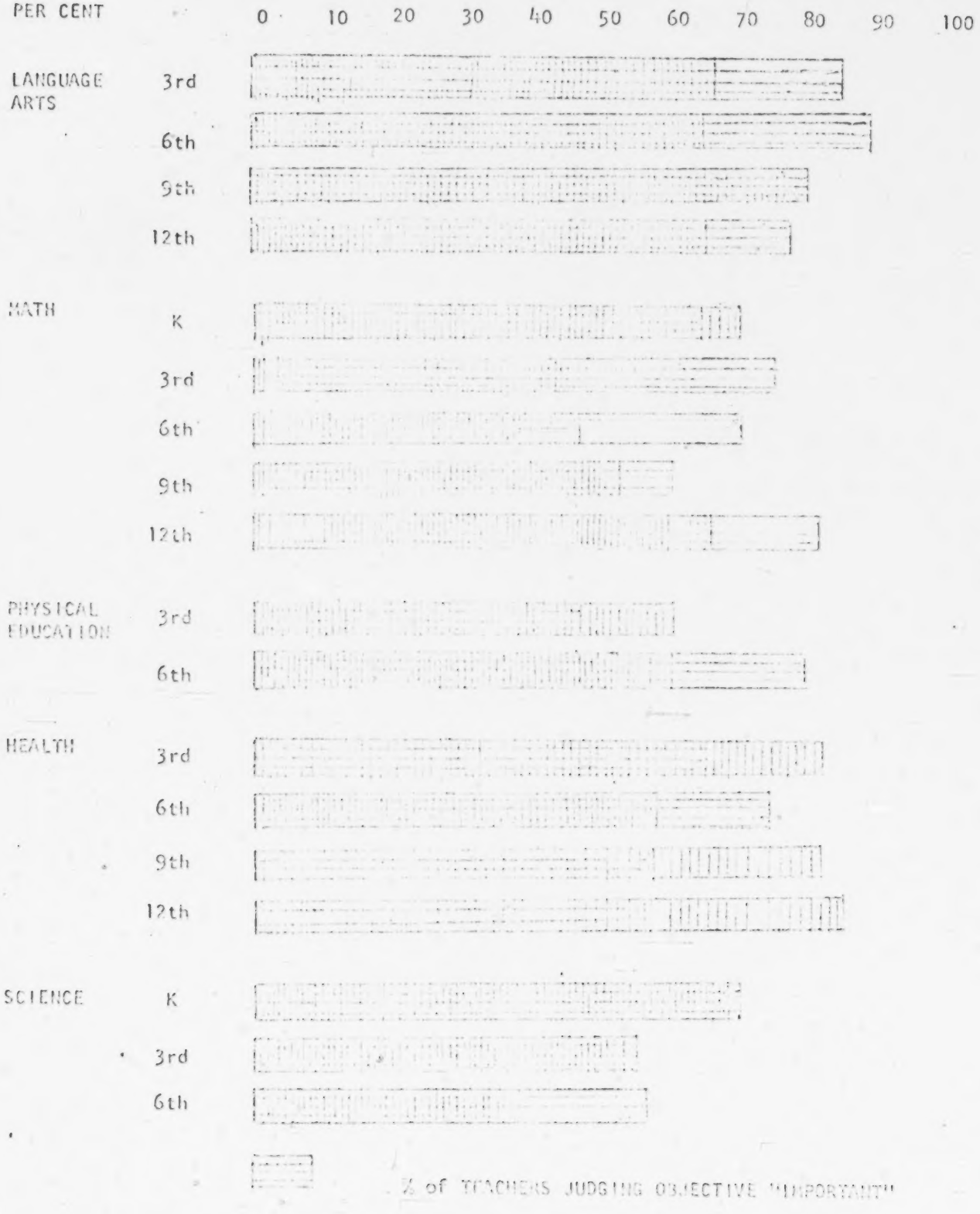
Eleventh - Grade Common Status Measure and Twelfth - Grade Subject Spearman Rank Order Correlation (Rho)*

		Language Arts	Science	Physical Educ.	Health	Music	Math	CSM-V	CSM-OCC
Basic Verbal Status	Rho	.28	---	.15	.25	.46	.22	---	.59
	N	29	---	18	23	19	26	---	32
Occupational Cognizance	Rho	.30	---	.05	.42	.45	.35	.59	---
	N	29	---	18	23	19	26	32	---

*Ranking by school mean

GRAPH NO. 1

Responses to Objectives: Teacher Judgment, Pupil Achievement



achievement of objectives to teacher's acceptance of objectives.

CSM and objective-referenced correlations. Estimated correlations (Spearman rhos based on school means) between the CSM and objective-referenced achievement item pools are given in Tables 12 and 13. These tables should be interpreted with extreme caution. The rhos are influenced by varying reliability of the achievement item pools, the difference among age groups in adjacent grades, and other statistical sources of imprecision.

Teacher judgment and pupil achievement. Responses to objectives were reported for two operations: (1) the percentage of teachers judging the objective to be important and (2) the percentage of correct pupil responses. These comparisons are shown on Graph #1 following Tables 12 and 13.

Summary

The Colorado Evaluation Project encompassed a wide range of activities and required the participation of students, teachers, classroom teachers, school and district administrators, personnel from the Colorado Department of Education and the United States Office of Education, and educational consultants and specialists. Each group played important functions in this project.

The purpose of the Colorado Evaluation Project was to field-test the Common Status Measures at grades four and eleven in conjunction with a state-wide assessment program based on objective-referenced testing instruments developed by the Colorado Department of Education for grades kindergarten, three, six, nine, and twelve. The evaluation was designed to determine educational needs and educational program performance in the state of Colorado.

A distinguishing feature of the Colorado Evaluation Project was its use of multiple-matrix sampling techniques to obtain reliable group data. Items were sampled from item pools for each subject at each grade level and randomly assigned to test forms. Districts, schools, classes and pupils were also randomly sampled. These techniques permit efficient and economical educational program assessment.

Results of several types of analyses are presented:

1. For each item and item pool, performance statistics were listed for the total pupil sample and by various sub-categories including: school program participation, urbanism, sex, ethnic background, whether a second language is spoken in the home, and socio-economic status.

2. Statistical operations were performed to obtain a profile of those students who performed in the lower 20% on each subject tested for a given grade level.

3. Performance statistics were also listed for each item and item pool for each school and school district in the sample.

4. For the Common Status Measures, comparisons between national data gathered during the Colorado Evaluation Project were performed.

5. Item quality, objective appropriateness and student performance were discussed for each objective-referenced item pool.

6. For each subject at each grade level, item intercorrelations were obtained.

7. Subject by subject correlations within each grade level and Common Status Measures by subject correlations for adjacent grade levels were also obtained.

8. Two video tapes, one dealing with occupational cognizance at the eleventh grade and one dealing with language arts at the third grade were produced. These tapes included interviews with students, teachers, and administrators, testing activity during the Colorado Evaluation Project, classroom activity, and discussions of the evaluation results.

CONCLUSIONS

The confluence of scientific and political thought has produced a new era in determination of educational need. From the activities and findings of the pilot project in assessment and evaluation these tentative conclusions were reached.

1. Educational need is demonstrated in relation to certain pupil populations. This need may be for pupils to master certain objectives, or the need may be for educators to form objectives more appropriate for the student population so identified. The need appears to be common in all subject areas represented by the tests.

2. Discrepancies exist in Colorado between what teachers expect of their pupils and what the pupils across the state can do. Each item of each test provides a fund of information as to specific areas of weakness. Interpretations of discrepancy information will be derived both from the objective itself and from pupil performance. The overriding concern is to relieve the educational need indicated by the discrepancy between the objective and its achievement.

3. Procedures of assessment appear adequate to measure statewide objective achievement of common learnings. That assessment system would have several innovative aspects:

   a. Pupils, teachers, and administrators would lose minimal instructional time to evaluation procedures.

   b. The system will maintain determination of the criteria for educational success in the hands of Colorado educators.

   c. The test questions may be revised, in part or as a whole, improved, and updated, following each testing cycle.

d. The system is more economically efficient than traditional testing programs.

e. The system insures that the evaluative data will be relevant to the planning needs of educational administrators.

f. In addition to giving educational decision-makers the information they require, the system provides a vehicle for educators to refine and explicate their educational objectives.

RECOMMENDATIONS FOR ACCOUNTABILITY

On the basis of the "findings" and "conclusions" discussed above, recommendations are offered for increasing accountability through assessment and evaluation. These recommendations form a model for a continuing assessment and evaluation program in Colorado.

Stating Objectives

Objectives should be reduced in number and increased in clarity. The objectives should be reviewed for cost estimates based upon current practice and modifications thereof by use of established and emerging organizational and technological advances. Probabilities for achieving these objectives should be computed statistically; such information can be utilized both in estimating costs and curriculum planning.

To realize these recommendations the following activities are suggested: (1) statement of objectives by inter-disciplinary committees and review by educational and community leaders. (2) collection and cataloging of objectives along with abstracts as to how to achieve these objectives and cost estimates.

Developing Test Items

Before submission to a statewide sample, all test items should be tried out in classrooms at the convenience of those involved. Items presently being utilized should be subject to further refinement. New

items assessing objectives in other domains and realms of learning should be developed.

Testing

Multi-matrix sampling should be continued as a basic scientific procedure in state assessment and evaluation. It produces needed information on student populations without the cost of every pupil taking all the tests. The procedures are not to be violated if statistical properties of the data are to be maintained; close monitoring of testing should be planned and acceptable by those people involved.

Analysis

Audiences to whom results are meaningful should be identified so that the reports can be planned and disseminated. Relationship of findings from the testing should be related to program inputs of resources and school processes. Inquiry should continue as to what input variables relate to educational quality and equal opportunity.

Recycling

Information as to student achievement, school program and processes and costs should be rechecked for reliability. Logical and practical expansions of this assessment system should be planned and executed, upon administrative and legislative authorization and support, by the Assessment and Evaluation Unit of the Colorado Department of Education.

Related Documents

Colorado Department of Education, Goals For Education In Colorado, The Department, Denver, Colorado, 1972

Heath, Robert W., Final Report, Comprehensive Evaluation Project, United States Office of Education, Washington, D. C., December, 1969

Olson, Arthur R., et al, Interim Report, Colorado Evaluation Project, Colorado Department of Education, Denver, Colorado, June, 1970

Pacific Educational Evaluation Systems, Colorado Evaluation Project: Statistical Tables and Technical Report, Colorado Department of Education, Denver, Colorado, August, 1970

Theoretical Works

Bloom, Benjamin (ed.). Taxonomy of Educational Objectives, The Classification of Educational Goals. Handbook I: Cognitive Domain, New York: David McKay Company, Inc., 1956

Lord, F. M. and M. Novick. Statistical Theory of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968

APPENDIX A

PUPIL CHARACTERISTIC ITEMS

ED 080571

T1 003 087

Elementary School Pupil Characteristic Items

Name of Student: _____

School: _____

Name of Teacher: _____

Date: _____

Time of Day: _____

To the proctor: Please answer the following questions with the aid of the student and/or school personnel. The aid of school personnel in answering item number four may be especially needed. Pupil's records should be checked to verify his responses.

1. What is this pupil's sex?

( ) Male ( ) Female

2. Indicate below if this pupil is a member of any of the following racial or national origin groups?

( ) American Indian
( ) Negro
( ) Oriental
( ) Spanish-surnamed American (Persons of Cuban, Mexican, or Puerto Rican descent)

( ) None of those listed

3. Is a language, other than English, regularly spoken in the pupil's home?

( ) Yes ( ) No

4. In the box below, please write the usual occupation of the person who is the primary supporter of this pupil's family. If you don't know, write "Don't Know" in the box below.

Please indicate below the most appropriate option describing the occupation you have written in the box above.

( ) Farm worker
( ) Farm manager or owner
( ) Unskilled worker, laborer, or domestic worker
( ) Semi-skilled worker
( ) Skilled worker
( ) Sales agents and representatives
( ) Technical
( ) Manager or foreman
( ) Official
( ) Professional
( ) Don't know

ED 080572

APPENDIX B

CONTENT AREAS

Sampled by Objectives-Referenced Items

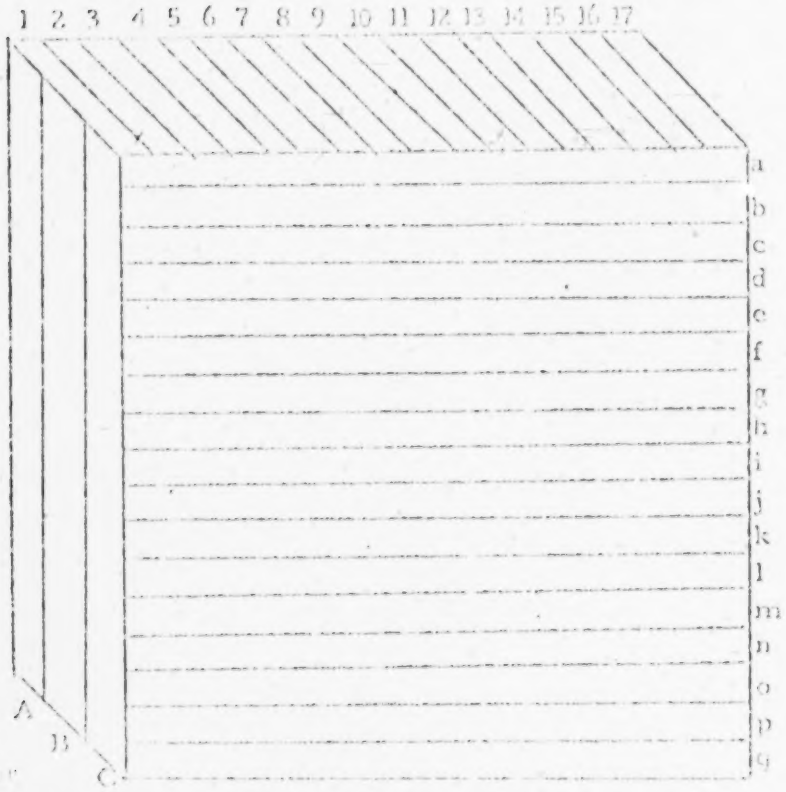
TM 003 088

Figure 1

OVERVIEW OF VARIABLES

Basic Educational Outcomes

1. Knowledge of Science info, skills, concepts.
2. Knowledge of Mathematics info, skills, concepts.
3. Knowledge of Social Studies info, skills, concepts.
4. Ability to read rapidly with comprehension.
5. Ability to communicate in Writing.
6. Ability to Spell correctly.
7. Knowledge and appreciation of Music.
8. Ability to draw and to appreciate Art.
9. Knowledge of Health skills and concepts.
10. Physical proficiency. (Already being measured)
11. Proficiency in and knowledge of Home Economics.
12. Proficiency in and knowledge of one or more Industrial Arts.
13. Cognizance of Occupational Opportunities.
14. Interest in School, School Subjects, and Education.
15. Value of Self, Family, Society.
16. Ability and initiative to solve real and pressing problems.
17. Ability and desire to participate in group work.



Program

Curricular:

a. Science
b. Math
c. Social Studies
d. Reading
e. Eng. Comp., Grammar, Lit.
f. Spelling
g. Music
h. Art
i. Health, Safety, Recreation
j. Physical Education
k. Home Economics
l. Voc. Ed/ Industrial Arts
m. Foreign Languages

Co-curricular:

n. Guidance
o. Health
p. Psychological
q. Social
r. Library
s. Food
t. Transportation

tudent
behavior
. Cognitive
. Affective
. Psychomotor

In the Context of--

Student and family information:

Age/Sex/Race or Ethnic Origin/Enrolled in this school 3 years ago/ Second language when in home/ I.Q. /Occupation of head of household/Education of mother.